On Tour: Harnessing Social Tourism Data for City and Point of Interest Recommendation

Tom Bewley[1] and Iván Palomares Carrascosa[2]

Department of Engineering Mathematics[1] and Department of Computer Science[2], University of Bristol, Bristol, UNITED KINGDOM

{ *tom.bewley, i.palomares* } *@bristol.ac.uk*

ABSTRACT

We introduce various data-driven models for recommending both city destinations and within-city points of interest to tourists. The models are implemented with a novel dataset of travel histories, derived from social media data, which is larger by size and scope than in prior work. All proposed models outperform simple baselines in cross-validation experiments, with the strongest variants reliably including tourists' true movements among their top recommendations.

1. INTRODUCTION AND BACKGROUND

Tourism is a popular activity for millions of people worldwide, with city breaks being a common form of holiday. Travellers face complex choices between potential destinations, then between specific points of interest (POIs) to visit upon arrival. In the existing literature, recommender systems have been developed to suggest both high-level destinations [4,6,7] and specific POIs [2,5,10,11] to individual tourists, informed by their explicitly- or implicitly-defined preferences and past visitors' experiences. In some works [1,8], the problem has been expanded to consider tourist groups with diverse preferences, which much be aggregated into a unified preference model or a list of 'compromise' recommendations that balances their needs.

Several recent tourism recommenders [9,10,13] operate by exploiting trends in large social datasets, such as those provided by the *Flickr* photo sharing platform. However, the reported results have been limited to only a small number of target cities, and no attempt has yet been made to perform both city- and POI-level recommendation using the same underlying dataset. In addition, relatively little effort has been dedicated to developing robust techniques for inferring accurate travel histories from the raw social data, and the accuracy of the derived datasets has not been questioned or investigated.

The overall aim of this project is to develop a set of recommendation tools to assist groups of travellers in the planning of city holidays, using both personalised data and population-wide statistics from a novel dataset of travel histories. This dataset is constructed by combining data from *OpenStreetMap* (OSM) with that from *YFCC100M* [12], the world's largest publicly-available media collection, which features accurate time and location tags for tens of millions of photos uploaded to *Flickr*. Two specific problems are addressed: city recommendation for groups of travellers, and within-city POI recommendation for a single traveller. In both cases, the output is a ranking of all available options, from most- to least-recommended.

2. RECOMMENDATION PROBLEMS AND MODELS

2.1 *City Recommendation*

The first problem addressed in this project is that of recommending new cities for groups of tourists to visit. Fig. 1 (left) illustrates the key elements of the city recommendation problem. T is a set of tourists, C is a set of cities and X is a set of POI categories. Each $t \in T$ has visited cities $V^t = \{c_1, \dots, c_l\} \subseteq C$, and a vector of city enjoyment values $e^t = [e_c^t \forall c \in C]$, which is non-zero only for the cities in V^t. t also has a vector of preferences over POI categories (e.g. churches, parks) $p^t = [p_x^t \forall x \in X]$, which sums to 1. Each city c has a distribution $i^c = [i_x^c \forall x \in X]$, which also sums to 1 and can be interpreted as the *importance* of each category in that city (e.g. churches are highly prevalent in Rome; parks comprise a large part of Vienna). A tourist group G is a subset of T which is the target of recommendation. The goal is to produce a ranking R of all unvisited cities $U^G = \bigcap_{t \in G} U^t$ (where $U^t = C - V^t$), ordered from most- to least-recommended for G. Considering first the single-tourist case ($G = \{t\}$) we propose three ranking models:

- *Tourist-tourist similarity* (TT). Compute the Jensen-Shannon distance (JSD) between e^t and $e^{t'}$ for each $t' \neq t$ and collect the 50 most similar tourists into a neighbourhood N^t. For each $c \in U^t$, define the recommendation score $S_{TT}^t(c)$ as the mean of $e_c^{t'} \cdot \left(1 - JSD\left(e^t, e^{t'}\right)\right)$ across all $t' \in N^t$. Produce the ranking R by sorting the cities in U^t by score. This model follows the conventional rationale of collaborative filtering: similarly-travelled tourists should continue to enjoy similar places in future.

Proc. 1st International 'Alan Turing' Conference on Decision Support and Recommender Systems (DSRS-Turing 2019)
The Alan Turing Institute, London, United Kingdom, 21-22nd November 2019
 ISBN 978-1-5262-0820-0

The Alan Turing Institute

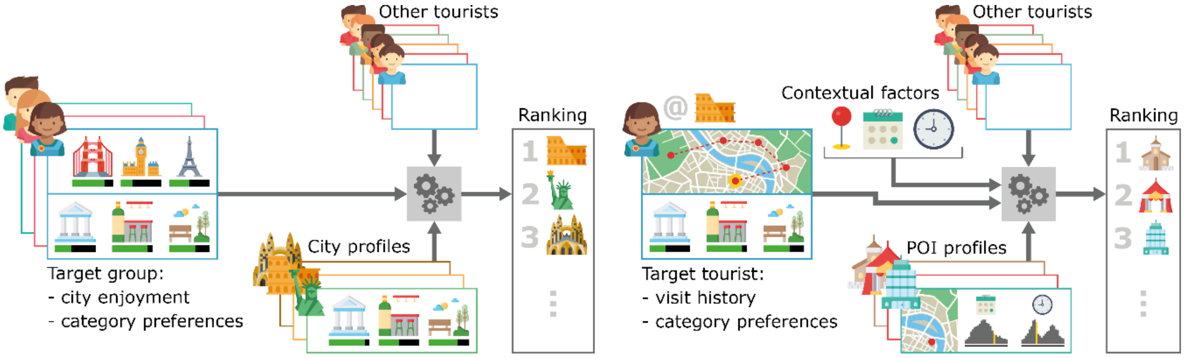

Figure 1. *Key elements of the city (left) and POI (right) recommendation problems.*

- *City-city similarity* (CC). For each $c \in U^t$, compute its similarity to each visited city $c' \in V^t$ in terms of their POI category importance distributions i^c and $i^{c'}$, again using JSD. Define the recommendation score for c as the mean of $e^t_{c'} \cdot \left(1 - JSD\left(i^c, i^{c'}\right)\right)$ across all $c' \in V^t$, and again rank by score. The rationale for this model is that tourists tend to visit cities with features in common.
- *Tourist-city similarity* (TC). For each $c \in U^t$, simply define the score for ranking as $1 - JSD(p^t, i^c)$. This model most closely resembles a conventional content-based recommender system: cities are recommended that align closely with the tourist's category preferences.

Multi-tourist groups ($|G| > 1$) add complexity since multiple preferences and visitation histories must be considered. We consider two techniques for producing aggregate recommendations for a group:

- *Aggregation-of-scores* (AoS). Complete scoring and ranking independently for each $t \in G$, then take the mean value of the per-tourist scores for each city.
- *Aggregation-of-preferences* (AoP). Define $V^G = \cup_{t \in G} V^t$ and p^G as the elementwise mean of the category preference vectors of the tourists in the group. Complete scoring and ranking with this group-level information, as if it represented a single tourist.

2.2 POI Recommendation

The second problem addressed is that of recommending POIs to visit during an ongoing trip to a specific city. Only the single-tourist case is considered due to project time constraints and foreseeable difficulties with group evaluation. Fig. 1 (right) illustrates the key elements of this problem. In addition to the notation introduced above, $P^c = \{p_1, \dots, p_o\}$ is the set of POIs in city c and $H^{t,c} = (v_1, \dots, v_q)$ is the chronologically-ordered history of *visits* by tourist t in city c, each defined by its start time, end time and associated POI. When finding a POI for t to visit next in c, the goal is to produce a ranking R containing all POIs in P^c, ordered from most- to least-recommended. The POI of the most recent visit $H^{t,c}_q$ is taken to be t's current location. We propose a hybrid model that uses the following six features to quantify the suitability of each POI p in this circumstance:

- $fPop$. The overall popularity of p, defined as $1 - 2^{-visitors/\alpha}$ where $visitors$ is the number of tourists that have been to p. Formally, $visitors = |\{t \in T :\ p \in H^{t,c}\}|$. In our implementation, $\alpha = 100$.
- $fCat$. t*'s* expected preference for p, as represented by p^t_x where x is the category of p.
- $fProx$. A measure of proximity / travelling convenience, defined as $1 - 2^{d/\beta}$ where d is the Euclidian distance in metres to p from t's current location. In our implementation, $\beta = 2000$.
- $fTime$. The appropriateness of p given the present hour of the day. This is computed by finding the proportion of recorded visits that lie within this hour, both to p itself (h_p), and also to p's category x at large across all cities (h_x). The two proportions are weighted by the number of visits to p, $visits$, using the formula $2^\varepsilon \cdot h_x + (1 - 2^\varepsilon) \cdot h_p$ where $\varepsilon = visits/\gamma$. In our implementation, $\gamma = 100$.
- $fDate$. The appropriateness of p given the present month, defined in an analogous fashion to $fTime$.
- $fHist$. A measure of the coincidence of p and the previously-visited POIs in $H^{t,c}$, within the visit histories of all other tourists $t' \neq t$. For each $p' \in H^{t,c}$, count the number of times p and p' both occur in another tourist's history, and weight each coincidence according to the time in hours Δ between the two visits, using $w = \zeta + (1 - \zeta)\exp(-\Delta / \kappa)$. The weighted values are summed across all p' and t', and normalised by p's total visit count, $visits$. In our implementation, $\zeta = 0.1$ and $\kappa = 24$.

We consider various techniques for mapping the six features into a single recommendation score for ranking: a trivial summing operation Sum, a linear regression model Lin and two small neural network topologies NN_3 (one 3-neuron hidden layer) and $NN_{6,6}$ (two 6-neuron hidden layers), both with logistic activation functions. In each case, we first z-normalise the features across a large bank of recommendation scenarios, since this improves learning stability.

3. TRAVEL HISTORIES DATASET

We synthesise a novel dataset of travel histories, consisting of approximately $812{,}000$ POI-level visits by $65{,}000$ tourists across 200 cities worldwide, using location-tagged photos from $YFCC100M$ [12]. We group the raw photos first by city (this information is available in the *Places* expansion pack) then by user, and sort chronologically. Consecutive photos taken within a 10 metre radius and 1 hour timeframe are combined into a single *visit*. These values are rather conservative, chosen so that bursts of photos taken at exactly the same location are grouped, but visits to two POIs on the same street or town square are not.

For each visit, we assemble a set of *visit words* from the user-provided titles, descriptions and tags, which can be compared with metadata for nearby POIs on OSM. Where a sufficiently strong match is found, the visit is labelled with the POI. Each POI can easily be assigned a category based on OSM's taxonomy of entity types. Consecutive visits to the same POI on the same day are combined into one, thereby preventing the aforementioned conservative visit creation parameters from yielding erroneous duplicate visits.

In a second 'bootstrapping' pass through the data, we use the visit words for already-labelled visits to assemble a distribution of word frequencies for each POI (e.g. *game* might be a high-frequency word for a stadium, and *lion* a high-frequency word for a zoo). The words for unlabelled visits are reassessed with respect to these distributions, and those that pass a fixed total summed probability threshold for a nearby POI are labelled with that POI. The bootstrapping pass boosts the number of labelled visits by approximately 30%.

A manual assessment of the dataset (viewing the underlying photos for one visit per city and looking for the presence of the POI) indicates that labelling accuracy is on the order of 75-85%, depending on how minor errors (e.g. labelling with the wrong building of the correct University) are penalised. This compares favourably with a dataset from prior work [9], which only uses location data for POI labelling, and for which we estimate an accuracy value around 45-60%. The 200-city coverage also far exceeds the eight included in the prior dataset. The statistics of our dataset align well with lists of popular POIs according to *TripAdvisor*, and reflect intuitive trends in the diurnal and seasonal variations in per-category POI visitation (e.g. restaurants are popular at mealtimes, and gardens in spring).

This dataset has the potential to serve as a widely-applicable resource of real-world tourist behaviour, complete with a measure of POI enjoyment through the proxy of number of photos taken (though it does lack an indicator of negative opinion, which may be sought at a later date through sentiment analysis of user-provided text). In light of its wide applicability, and its derivation from public-domain resources, We have published the dataset at https://github.com/tombewley/OnTour-TourismRecommendation/tree/master/dataset.

It exists as a single 65MB JSON file whose core elements are individual visits. Each visit entry contains the POI (as represented by the unique OpenStreetMap identifier), the start and end timestamps and the number of photos taken. Visits are categorised at the highest level by user, then by city, and ordered chronologically. The dataset also includes two separate subschemas containing the total number of photos taken by each user in each city, and the name, category and coordinates of each POI.

4. IMPLEMENTATION, EVALUATION AND OPTIMISATION

We implement our models in Python to work with the travel histories dataset. Preferences are defined in terms of photo counts and metadata: city enjoyments e^t and category preferences p^t are defined by the fraction of t's photos taken in each $c \in C$ or $x \in X$. Similarly, per-city POI category importance values i^c are the fraction of photos taken at each category. Tourist t's history $H^{t,c}$ is defined directly as their visits in c within the dataset.

Following previous work [9,13], our evaluation method is one of cross-validation, which takes the general form of *forgetting* an item from the dataset and assessing the models' ability to predict it using what remains. For city recommendation: a group of tourists is assembled who have all visited a common city c^* (as well as at least two others), this city is removed from their histories and the remaining data used as input to the models. For POI recommendation (single-tourist only): a tourist's history in one city is cut short at a point that ensures at least 20 visits remain, and the time gap between the visits immediately before and after (v and v^*) is at most 8 hours. The POI of v is taken as the tourist's current location and that of v^* is the forgotten POI p^*. In both city and POI recommendation, performance is measured via r^*, the position of c^*or p^* in the ranking. For the POI recommendation problem, this is reported as a fraction of the number of POIs in the city.

The Alan Turing Institute

For the POI recommender, the scoring methods Lin and NN_3 and $NN_{6,6}$ need to be trained on the dataset. To do this, we create a bank of all valid recommendation scenarios (15,326 in total), partition into training, validation and test sets (60: 20: 20 split) and run the training set through the model. In each scenario, we compute r^*, sample a random POI placed higher in the ranking, and define the error as the difference between the two scores. We employ a stochastic gradient descent optimiser, and perform a secondary set of weight updates to non-aggressively push the score magnitudes into the range $[0,1]$ for stability purposes. Training is stopped when validation set performance begins to decrease, which prevents overfitting.

5. RESULTS

5.1 City Recommendation

The box plots in fig. 2 summarise the performance of the three models on 500-case test sets with a group size $|G|$ of 1, 3 and 10. Cases are sampled uniformly across cities to avoid biasing to popular destinations. A clear result is that all three models perform markedly better than random. However, the collaborative filtering model TT is the strongest by a wide margin, with a median r^* of 8 out of 200 with $|G| = 1$. Performance improves as the group size increases. The best results are attained with $|G| = 10$ using AoS; here the r^* quartiles lie at 1, 1 and 6, which implies that in fully half of the test cases, the model ranks the forgotten city in first place.

With $|G| = 1$, the median r^* for the CC and TC models is 61 and 60 respectively; the two are almost inseparable. A gap does open up for larger group sizes, with TC ranking the forgotten city around 20 places higher on average for $|G| = 10$, most notably when the AoS aggregation method is used. With the other two models, the aggregation method has no consistent effect. An investigation of the weaker performance of the CC and TC models reveals few strong results, though it is clear that more visits to the forgotten city afford a more accurate estimation of its POI category importance vector, and in turn a higher ranking.

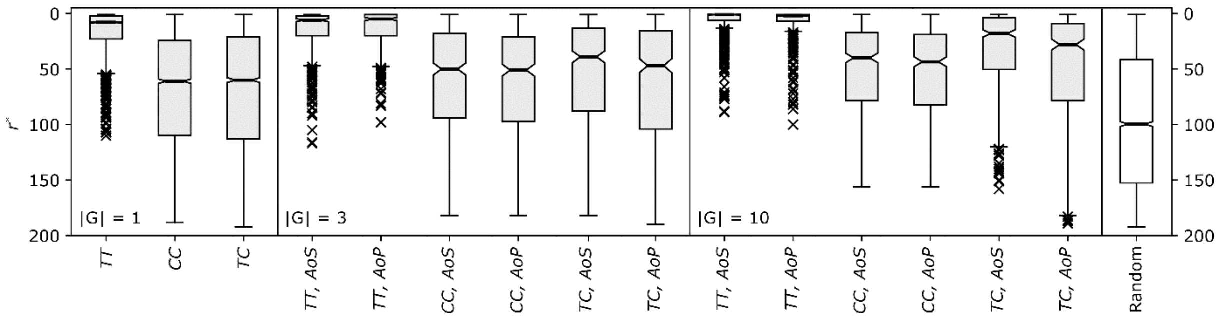

Figure 2. *City recommendation performance with each model and various group sizes.*

5.2 POI Recommendation

Fig. 3 shows the performance of our model with each scoring method on the 3,055-sample test set, alongside baseline results from using each feature in isolation, as well as random ranking. All methods outperform all baselines, and are remarkably similar despite differences in complexity. This suggests that even a trivial summation of features can reach close to the maximum attainable performance on the dataset, given the set of features. That said, the larger neural network topology ($NN_{6,6}$) is the strongest, placing the forgotten POI at a median position of 6% down the ranking of candidate POIs. The other methods perform around 1-2% worse by rank percentage. On a case-by-case level, $NN_{6,6}$ gives the highest or joint-highest r^* 58% of the time.

In surrounding analyses, we find that visits in popular and Anglophone cities are predicted best, likely due to better data quality in these locations. Static POI categories (e.g. monuments) are also better predicted than those that are weather- or event-sensitive (e.g. theatres). Correlation analysis of the four scoring methods shows a high degree of similarity in their r^* values across the test set; if one method performs well, the others are likely to do the same. Further correlation analysis of each method with respect to each feature shows that the learning models place greater weight on $fProx$, $fPop$ and $fHist$ than the other three. This aligns with the greater predictive efficacy of the baseline rankings generated by these features. A final set of feature-to-feature correlations shows that they are largely independent, which is desirable since this maximises informativeness.

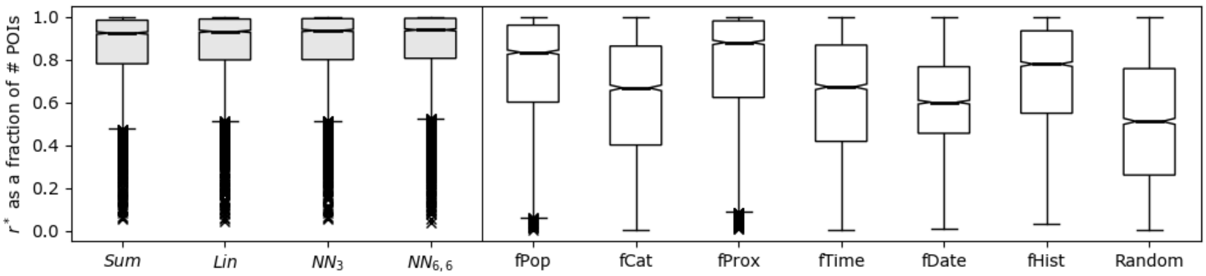

Figure 3. *POI recommendation performance with each scoring method, and single-feature and random baselines.*

6. CONCLUSION

We have presented models for the data-driven recommendation city destinations and POIs and evaluated them by cross-validation on a novel dataset. In all cases, performance exceeds that of simple baselines. For city recommendation, the collaborative filtering model TT performs strongest by a wide margin. For POI recommendation, even a simple summation of features is effective, though small gains can be attained by using a neural network for scoring. Future work could involve model refinement (new features and optimised parameters), testing via user trials, and deployment as an interactive web application.

Distinct from our specific choice of recommender models, the travel histories dataset itself provides a large (812,000 entries; 65,000 tourists; 200 cities) and high-accuracy resource of POI-level touristic visits. It is synthesised from two freely-available data sources – YFCC100M and OpenStreetMap – thus can itself can be freely used as part of future work in the field of tourism recommendation.

More details on this project can be found in the full MSc thesis paper [3], which is currently being graded but will be made available at https://tombewley.com/msc.

Acknowledgements: Many thanks to Ercan Ezin, who was co-supervisor for this MSc thesis project.

7. REFERENCES

[1] L Ardissono et al. (2003), Intrigue: personalized recommendation of tourist attractions for desktop and hand held devices, *Applied artificial intelligence 17.8-9: 687-714.*

[2] M Batet et al. (2012), Turist@: Agent-based personalised recommendation of tourist activities, *Expert Systems with Applications 39.8: 7319-7329.*

[3] T Bewley (2019), On Tour: Harnessing Social Tourism Data for City and Point-of-Interest Recommendation, *MSc Thesis, University of Bristol.*

[4] L Cao et al. (2010), A worldwide tourism recommendation system based on geotagged web photos, *IEEE International Conference on Acoustics, Speech and Signal Processing. IEEE.*

[5] L Castillo et al. (2008), Samap: An user-oriented adaptive system for planning tourist visits, *Expert Systems with Applications 34.2: 1318-1332.*

[6] J A Delgado and R Davidson (2002), Knowledge bases and user profiling in travel and hospitality recommender systems, *Proceedings of the Enter 2002 Conference. Springer Verlag, Wien, NY, 1–16.*

[7] M Goossen et al. (2009), My ideal tourism destination: Personalized destination recommendation system combining individual preferences and GIS data, *Information Technology & Tourism 11.1: 17-30.*

[8] A Jameson (2004), More than the sum of its members: challenges for group recommender systems, *Proceedings of the working conference on Advanced visual interfaces. ACM.*

[9] K H Lim et al. (2015), Personalized tour recommendation based on user interests and points of interest visit durations, *Twenty-Fourth International Joint Conference on Artificial Intelligence*.

[10] K H Lim et al. (2017), Personalized itinerary recommendation with queuing time awareness, *Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval.*

[11] R Logesh et al. (2019), Efficient user profiling based intelligent travel recommender system for individual and group of users, *Mobile Networks and Applications 24.3: 1018-1033.*

[12] B Thomee et al. (2015), YFCC100M: The new data in multimedia research, *arXiv:1503.01817*.

[13] L Wan et al. (2018), A hybrid ensemble learning method for tourist route recommendations based on geo-tagged social networks, *International Journal of Geographical Information Science 32.11: 2225-2246.*

 ISBN 978-1-5262-0820-0